



An Imbalance-Aware Feature-Optimized Heterogeneous Ensemble for Robust Network Intrusion Detection

Dr. Deelip Ananda Patil^{*1}, Mr. Ashish Dhananjay Kulkarni², Mr. Atul Manoj Otari³, Mr. Ratnesh Vikas Patil⁴, Miss. Thorat Naina Shivaji⁵, Miss. Suryawanshi Shital Jyotiram⁶

¹Department of Computer Applications, Rajarambapu Institute of Technology Ishwarpur, India. E-mail: deelip.patil@ritindia.edu

²Department of Computer Applications, Rajarambapu Institute of Technology Ishwarpur, India. E-mail: ashish.kulkarni@ritindia.edu

³Department of Computer Applications, Rajarambapu Institute of Technology Ishwarpur, India. E-mail: atul.otari@ritindia.edu

⁴Department of Computer Applications, Rajarambapu Institute of Technology Ishwarpur, India. E-mail: ratnesh.patil@ritindia.edu

⁵Department of Computer Applications, Sadguru Gadge Maharaj College, Karad, India. E-mail: thoratsnaina60@gmail.com

⁶Department of Computer Applications, Sadguru Gadge Maharaj College, Karad, India. E-mail: yanviswara4@gmail.com

Abstract

Accurately classifying types of cyberattacks continues to be a major challenge for network intrusion detection systems. This is especially true when dealing with dimensional traffic data, imbalanced classes and attacks that change over time. To tackle these issues this study introduces an ensemble framework that is aware of class imbalance and optimized for feature selection. The framework is built to handle intrusion detection and uses five base models. LightGBM, XGBoost, Random Forest, a Deep Neural Network (DNN) and an FT-Transformer. These models work together through voting to make the final prediction. The experiments use the CSE-CIC-IDS2018 dataset. Attack instances are grouped into four categories: BENIGN, DoS, DDoS and BruteForce. To ensure the results are reliable and to avoid information leakage several preprocessing steps are applied. These include chronological data splitting removing features using correlation-based filtering and Recursive Feature Elimination (RFE). The framework also uses train- standardization and applies custom methods to handle class imbalance for each model. As a result the original 78 traffic features are reduced to 30 that are most useful. The ensemble is tested using stratified five-fold out-of-fold (OOF) validation and an unseen chronological test set. This setup allows evaluation of both prediction consistency and how well the model performs on time-based data. The ensemble achieves an OOF accuracy of 95.38% and a Macro-F1 score of $95.22\% \pm 0.02\%$. On the chronological test set it reaches 95.23% accuracy and 95.09% Macro-F1. Performance is very strong for BENIGN and DDoS traffic with F1-scores of 99.98% and 99.99% respectively. For BruteForce the recall is 95.70% while DoS has a recall at 87.14%. Most of the mistakes occur between DoS and BruteForce showing the difficulty in distinguishing these two. The small difference between OOF and test set results shows that the framework is stable and can generalize over time. Overall the results show that this approach is both reliable and effective, for network intrusion detection.

Keywords: Network Intrusion Detection, CSE-CIC-IDS2018, Ensemble Learning, Feature Selection, RFE, Class Imbalance, LightGBM, XGBoost, FT-Transformer, Deep Learning.

1. Introduction

The rapid growth of networked systems, IoT devices, and cloud services has greatly expanded the range of attacks that bad actors can carry out, thereby making advanced cyberattacks a greater concern. In such a situation, Network Intrusion Detection Systems (NIDS) are essential for detecting harmful network activity. Nowadays, machine learning and deep learning techniques are commonly used in this area since they are able to learn patterns directly from network traffic. Yet there are still a number of problems that hinder the effectiveness of detection: unnecessary or duplicate features increase complexity without offering any benefit, the various types of attacks are not evenly distributed, and attacks that are similar to each other are often difficult to distinguish. Feature selection is useful for eliminating overlapping or redundant attributes, and ensemble learning enables different classifiers to combine their individual strengths. A great deal of research has been carried out on combining feature selection, methods for dealing with class imbalance, and ensemble techniques in order to improve the performance of NIDS [1] [2] [14] [15] [16] [18] [22].

A further issue is the way in which model evaluation is carried out. When the data is split randomly and preprocessing takes place before the split, the test data may leak into the training data, leading to overly optimistic results that do not represent the model's true performance in the real world. By splitting the data in a chronologically sequential manner, on the other hand, a more realistic test is obtained, as the model has to work with data that comes after the training period. In order to tackle these problems, this study presents an imbalance-aware, feature-optimized framework for robust network intrusion detection, which has been tested on the CSE-CIC-IDS2018 dataset.

The framework proposed reduces the initial number of 78 features to 30 that are considered useful by means of correlation-based feature filtering together with Recursive Feature Elimination (RFE). The five base models — LightGBM, XGBoost, Random Forest, a Deep Neural Network (DNN), and an FT-Transformer — are combined through soft probability voting, and special techniques for dealing with imbalance are used

in order to enhance the detection of the minority classes. Evaluation of the model is carried out using a leakage-aware, out-of-fold validation method, and it is additionally tested on a set that has not been seen chronologically to check how well it generalises over time.

The last heterogeneous ensemble attains an out-of-fold accuracy of 95.38% and a Macro-F1 score of 95.22%. When applied to the test set which was seen chronologically, it achieves an accuracy of 95.23% and a Macro-F1 of 95.09%. These results indicate a strong overall performance, together with the fact that the main cause of error is still the distinction between DoS and BruteForce traffic.

2. Literature Review

Network Intrusion Detection Systems (NIDS) have increasingly adopted machine learning (ML), deep learning (DL), feature selection, and ensemble learning to improve the detection of diverse cyberattacks. Recent research indicates that intrusion detection performance is strongly influenced by the dimensionality of network traffic, class imbalance, model selection, and the ability to generalize across different attack patterns. Consequently, current research has shifted from standalone classifiers toward integrated frameworks combining feature optimization, imbalance mitigation, and ensemble learning.

Feature selection is an important component of NIDS because network traffic datasets contain redundant, highly correlated, and irrelevant attributes. Abhale and Avulapalli [1] investigated Recursive Feature Elimination (RFE) with resampling for intrusion detection in Wireless Sensor Networks, demonstrating the usefulness of feature elimination in improving intrusion detection. Mhawi et al. [14] proposed advanced feature-selection-based hybrid ensemble learning algorithms for NIDS, emphasizing the role of optimized feature subsets in improving classification. Alsaffar et al. [2] combined hybrid feature selection with ensemble learning using mutual information and Boruta along with RF, XGBoost, CatBoost, and MLP classifiers. Similarly, Nassreddine et al. [16] investigated correlation and embedded feature-selection techniques with ensemble learning. Jayalatchumy et al. [11] employed an improved Crow Search-based feature-selection approach for IoT intrusion detection, while Ji et al. [12] investigated optimal feature selection for cyber-physical-system intrusion detection. These studies establish that reducing redundant features can improve model efficiency and classification performance.

Class imbalance is another major challenge in intrusion detection. Attack categories are often unevenly distributed, and models optimized primarily for overall accuracy may fail to adequately detect minority or difficult attack classes. Musthafa et al. [15] investigated balanced class distributions, feature selection, and ensemble machine learning for IoT intrusion detection. Chen et al. [8] proposed a hybrid IDS combining data resampling and deep learning. Nigar and Mustafa [17] integrated hybrid data resampling with feature optimization, while Talukder et al. [23] addressed large and imbalanced intrusion datasets through oversampling, stacking feature embedding, and feature extraction. Dasari et al. [9] incorporated Boruta feature selection into a stacked ensemble-based NIDS. These studies demonstrate the effectiveness of explicitly addressing class imbalance rather than relying only on conventional classification.

Recent research has also focused specifically on minority and rare attack detection. Oh et al. [18] proposed the Minority Class Highlighting Ensemble method to improve classification under class imbalance. Sălișteanu et al. [20] introduced a cost-sensitive stacking ensemble with hybrid feature selection for rare-attack detection, demonstrating the value of assigning greater importance to difficult attack classes. Mchina et al. [13] proposed adaptive class-aware feature selection for high-dimensional and imbalanced multiclass intrusion detection. Such approaches indicate a growing trend toward integrating imbalance awareness into feature selection and model learning.

Ensemble learning has emerged as an effective strategy for improving the robustness of NIDS. Instead of depending on a single classifier, ensemble methods combine multiple models to exploit their complementary strengths. Alsaffar et al. [2] employed a stacking framework involving multiple tree-based and neural models, while Badiger and Shyam [6] proposed a multiclass NIDS using stacking ensemble learning with hybrid feature selection. Dasari et al. [9] similarly developed a stacked ensemble optimized using Boruta feature selection. Yang et al. [25] proposed an optimized adaptive ensemble model with feature selection, highlighting the benefits of combining feature optimization and multiple learners. More recent studies have explored dynamic ensemble learning and adversarial augmentation [19], multi-metaheuristic feature selection with ensemble learning [7], and DeepSMOTE-driven stacking ensembles [22].

Deep learning has further expanded the capabilities of intrusion detection systems by enabling models to learn complex nonlinear relationships from network traffic. Ayantayo et al. [4] investigated feature fusion with deep learning for network intrusion detection, while Chen et al. [8] integrated resampling with deep learning. Recent research has also explored attention-guided architectures for rare-attack detection in industrial IoT environments [21]. These studies suggest that combining conventional ML models with DL architectures can provide complementary representations of network traffic.

Despite these advances, several limitations remain in existing research. Many studies rely on resampling techniques or additional meta-learners, which can increase computational complexity and modify the original data distribution. Furthermore, conventional random train-test partitioning may not adequately evaluate performance on temporally evolving network traffic. Feature-selection and preprocessing

procedures may also introduce information leakage if they are performed before data partitioning. In addition, reporting only overall accuracy can conceal poor performance for individual attack classes. To address these limitations, the proposed study develops an imbalance-aware feature-optimized heterogeneous ensemble combining LightGBM, XGBoost, Random Forest, DNN, and FT-Transformer using soft probability voting. Correlation-based feature filtering and RFE are used to reduce redundant features, while model-specific class and sample weighting addresses class imbalance without altering the original observations. The framework also employs leakage-aware out-of-fold evaluation and a chronological unseen test set to assess temporal generalization. Thus, the proposed approach integrates feature optimization, imbalance handling, heterogeneous learning, and realistic evaluation within a unified NIDS framework.

3. Research Methodology

The proposed methodology develops an imbalance-aware, feature-optimized soft-voting ensemble model for multiclass network intrusion detection using the CSE-CIC-IDS2018 dataset.

3.1 Dataset Preparation

The selected CSE-CIC-IDS2018 network traffic files undergo integration and data cleaning processes. This involves addressing missing values, infinite values, duplicate records, and irrelevant attributes. The original attack categories are consolidated into primary classes, namely BENIGN, BruteForce, DoS, and DDoS, to establish a consistent multiclass classification framework. Given the dataset's substantial number of benign instances relative to attack instances, benign-class capping is implemented to mitigate class dominance.

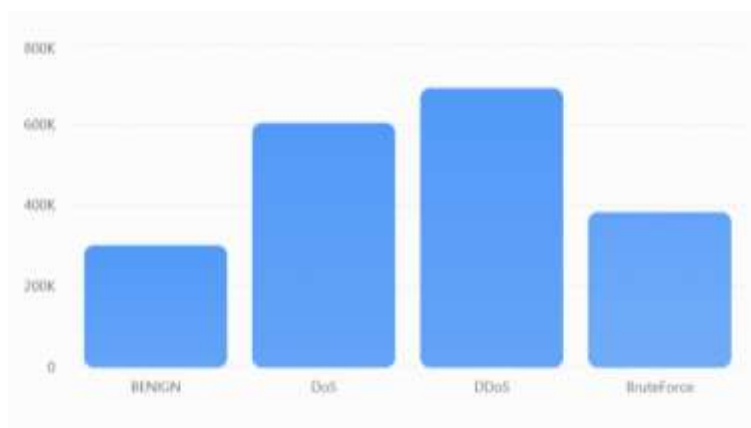


Fig 1: Class distribution of the CSE-CIC-IDS2018 dataset.

Furthermore, imbalance-aware techniques are employed during model training to enhance the recognition of minority classes. The cleaned dataset is partitioned chronologically into training, validation, and testing sets. This chronological partitioning, as opposed to a completely random split, ensures that the final model is evaluated on traffic data from a later time period, thereby providing a more realistic assessment of generalization.

3.2 Cross-Validation and OOF Prediction

The training data is subjected to a stratified five-fold cross-validation strategy. This approach generates Out-of-Fold (OOF) predictions, ensuring reliable predictions from models on samples excluded from their respective fold training.

3.3 Feature Optimization

Feature optimization is performed independently within each fold. First, correlation-based filtering removes highly redundant features. Subsequently, Recursive Feature Elimination (RFE) identifies the most informative features for intrusion classification.

3.4 Feature Transformation

The standardized optimization of the feature set is conducted prior to its provision to the learning models. This ensures a consistent numerical representation, thereby facilitating the effective training of various model architectures.

3.5 Base Learners

Five complementary learning models are employed:

- **LightGBM** – efficient gradient-boosting model for tabular network data.
- **XGBoost** – powerful tree-based boosting classifier.
- **Random Forest** – ensemble of decision trees providing robust nonlinear classification.
- **DNN** – learns complex nonlinear relationships between network features.
- **FT-Transformer** – transformer-based architecture designed for tabular feature learning.



Fig. 2: Imbalance-Aware Feature-Optimized Soft-Voting Ensemble for Intrusion Detection

3.6 Custom Soft-Voting Ensemble

The probability outputs produced by the five base learners are integrated through a custom weighted soft-voting mechanism. This ensemble approach, which does not depend on a single classifier, synthesizes the complementary predictions of all models to determine the final intrusion class.

3.7 Model Evaluation

The OOF predictions are first evaluated to assess the ensemble during the training stage. The final ensemble is subsequently trained using the **training and validation data** and evaluated on the **unseen chronological test set**.

Performance is measured using:

- Accuracy
- Precision
- Recall
- Macro-F1 Score
- Classification Report
- Confusion Matrix

4. Results and Discussion

The proposed imbalance-aware, feature-optimized soft-voting ensemble was evaluated on the CSE-CIC-IDS2018 dataset containing 1,969,879 samples and 78 features across BENIGN, DoS, DDoS, and BruteForce classes. A chronological train-validation-test split was used to evaluate the model on unseen future traffic.

4.1 Cross-Validation Results

The ensemble showed highly consistent performance across five stratified OOF folds.

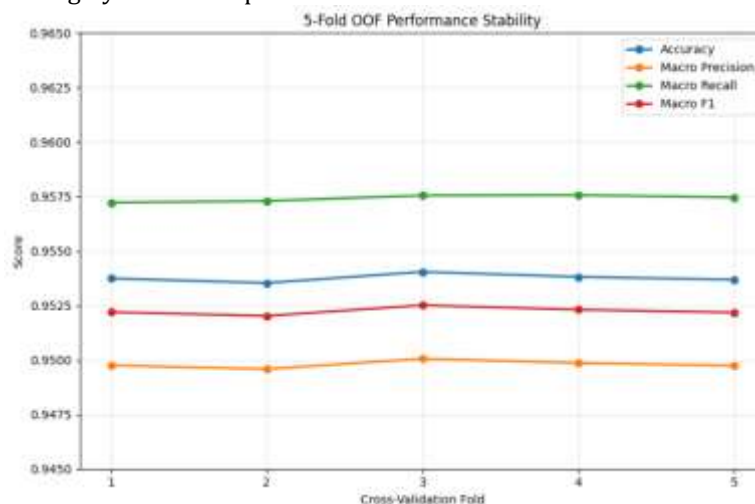


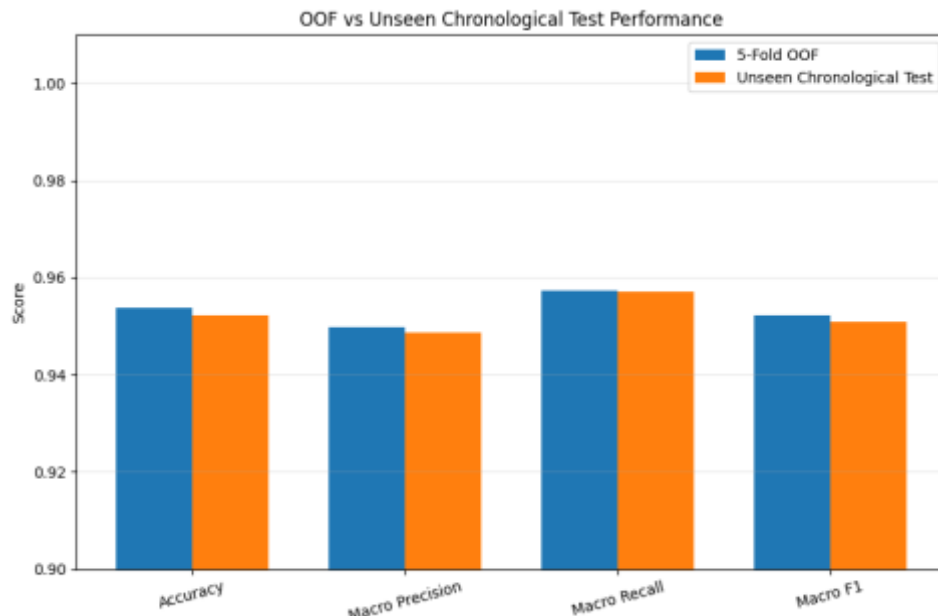
Fig. 3: 5-Fold OOF Performance Stability

The very small standard deviations indicate stable performance across different folds. Achieving a high Macro Recall is crucial for intrusion detection, as it indicates effective identification across various traffic classes.

4.2 Feature Optimization

The feature optimization process reduced the original 78 features to 30 selected features, representing approximately 61.5% dimensionality reduction. Despite this reduction, the ensemble maintained an OOF accuracy of 95.38%, indicating that the selected features retained sufficient discriminatory information while reducing feature complexity.

4.3 Unseen Test Performance

**Fig. 4: OOF vs Chronological Test**

All 393,978 unseen chronological test samples were used for the final ensemble. The test accuracy remained almost the same, dropping slightly from 95.38% OOF to 95.23%, with Macro F1 dropping from 95.22% to 95.09%. This is a small performance difference, which reflects good generalization to unseen network traffic in time.

4.4 Class-Wise Performance

The model worked well in the case of BENIGN and DDoS traffic. The F1-Score of BENIGN is 99.98%. The F1-Score of DDoS is ~99.99%. This shows an accurate detection of both traffic type and DDoS traffic. For DoS the model achieved 96.97% precision 87.14% recall and 91.79% F1-score. The lower recall is mainly due, to DoS samples being classified as BruteForce. The BruteForce class had a high recall of 95.70% but a relatively low precision of 82.49% and an F1-score of 88.60%.

4.5 Confusion Matrix Analysis

The primary misclassifications were between the DoS and BruteForce classes. Specifically, 15,481 DoS samples were misclassified as BruteForce and 3,275 BruteForce samples were misclassified as DoS. For the other traffic types, very few errors were observed for BENIGN and DDoS traffic.

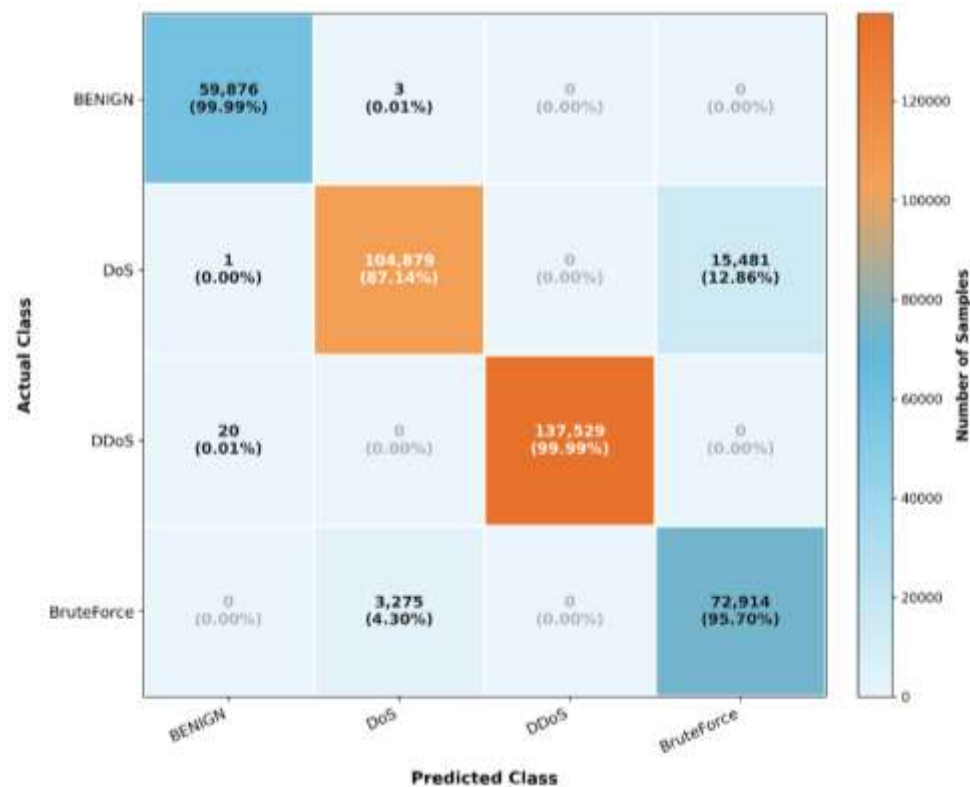


Fig. 5: Confusion Matrix on Unseen Chronological Test Data

From this pattern, it is clear that both BruteForce traffic and DoS traffic have similar properties, and a model that is able to distinguish between them is more challenging. More studies are required to gain better discrimination between these two types of attacks. Overall, the results indicate that the proposed approach is effective in detecting intrusions across different classes, including unseen network traffic in the temporal domain.

Conclusion

For the multi-class intrusion detection system (IDS) using the CSE-CIC-IDS2018 dataset, the soft-voting ensemble was proposed in this article, in which the feature-optimized and imbalance-aware soft-voting ensemble method was proposed to further enhance the performance of the multi-class intrusion detection system. The proposed framework is a fusion of five model architectures: LightGBM, XGBoost, Random Forest, DNN, and FT-Transformer and it reduces the feature space from 78 to 30.

The test set that was used was unseen chronological and the ensemble achieved 95.09% macro F1 Score, 95.70% macro recall Score, 94.86% macro precision Score, and 95.23% accuracy Score. The five-fold OOF is well correlated with the test result, and has good stability and temporal generalization. The model did very well in F1-scores for the BENIGN traffic with 99.98%, and the DDoS traffic with 99.99%. The most important difficulty, however, was that it was hard to distinguish between two attack classes, DoS and BruteForce, resulting in the most misclassified attacks. Future work could focus on feature engineering within the class level, more advanced methods for dealing with imbalances, improved thresholds optimization and improved temporal modelling to improve discrimination between these attack classes.

The experimental results for the whole experiment confirm that the proposed framework is an effective and efficient network intrusion detection approach that can be transferred to detect multiple classes of network intrusions.

References

1. Abhale, A. B., & Avulapalli, J. R. (2023). Enhancing intrusion detection recursive feature elimination with resampling in WSN. *International Journal of System Assurance Engineering and Management*, 14, 2642-2660. <https://doi.org/10.1007/s13198-023-02128-3>
2. Alsaffar, A. M., Nouri, M., & Zolbanin, H. (2024). Shielding networks: enhancing intrusion detection with hybrid feature selection and stack ensemble learning. *Journal of Big Data*, 11. <https://doi.org/10.1186/s40537-024-00994-7>
3. Aman, S., & Researcher, I. M. (2026). Robust Network Intrusion Detection Using Deep Stacking Ensembles and Hybrid Feature Selection. *Iraqi Journal of Communications and Computer Networks*. <https://doi.org/10.58564/ijccn.2.1.2026.14>

4. Ayantayo, A., Kaur, A., Kour, A., Schmoor, X., Shah, F., Vickers, I., Kearney, P., & Abdelsamea, M. M. (2023). Network intrusion detection using feature fusion with deep learning. *Journal of Big Data*, 10, 1-24. <https://doi.org/10.1186/s40537-023-00834-0>
5. B, G., Jadhav, V. D., Kashid, S. S., Newaskar, D., R, M., & S, S. (2025). Enhanced Intrusion Detection in SDN Using Ensemble Learning and Hybrid Feature Optimization. *2025 IEEE DELCON - International Conference on Recent Smart Technologies in Engineering for Sustainable Development*, 1-6. <https://doi.org/10.1109/delcon68055.2025.11400446>
6. Badiger, V. S., & Shyam, G. K. (2025). A Multiclass Network Intrusion Detection System Using Stacking Ensemble Technique with Hybrid Feature Selection. *Journal of Advances in Information Technology*. <https://doi.org/10.12720/jait.16.3.342-356>
7. Bagadi, S., Rao, G., & Vindhya, A. (2026). A Multi-Metaheuristic Feature Selection and Ensemble Learning Framework for Intrusion Detection Systems. *2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS)*, 144-150. <https://doi.org/10.1109/icicds70526.2026.11604694>
8. Chen, H., You, G., & Shiue, Y. (2024). Hybrid Intrusion Detection System Based on Data Resampling and Deep Learning. *International Journal of Advanced Computer Science and Applications*. <https://doi.org/10.14569/ijacsa.2024.0150214>
9. Dasari, A., Biswas, S., & Purkayastha, B. (2025). ASENIDSB: A Stacked Ensemble-based Network Intrusion Detection System Optimized with Boruta Feature Selection. *Journal of Intelligent & Fuzzy Systems*, 50, 1073 - 1092. <https://doi.org/10.1177/18758967251358591>
10. Elsheikh, Y., Dubey, C. A., Kowalski, E., Mahajan, S., Prasad, D., B., Konduru, R. K., Chandra, S., & K. (2026). AI-Powered Adaptive Network Defense Through Feature Optimization and Ensemble Cyber-Threat Recognition. *Journal of Intelligent Decision Making and Information Science*. <https://doi.org/10.59543/jidmis.v3.848>
11. Jayalatchumy, D., Ramalingam, R., Balakrishnan, A., Safran, M. S., & Alfarhood, S. (2024). Improved Crow Search-Based Feature Selection and Ensemble Learning for IoT Intrusion Detection. *IEEE Access*, 12, 33218-33235. <https://doi.org/10.1109/access.2024.3372859>
12. Ji, R., Kumar, N., & Padha, D. (2024). Hybrid Enhanced Intrusion Detection Frameworks for Cyber-Physical Systems via Optimal Features Selection. *Indian Journal Of Science And Technology*. <https://doi.org/10.17485/ijst/v17i30.1794>
13. Mchina, J. P., Mduma, N., & Sinde, R. (2026). Adaptive class-aware feature selection for high-dimensional and imbalanced multi-class network intrusion detection. *Frontiers in Big Data*, 9. <https://doi.org/10.3389/fdata.2026.1871346>
14. Mhaw, D. N., Aldallal, A., & Hassan, S. (2022). Advanced Feature-Selection-Based Hybrid Ensemble Learning Algorithms for Network Intrusion Detection Systems. *Symmetry*, 14, 1461. <https://doi.org/10.3390/sym14071461>
15. Musthafa, M., Huda, S., Koder, Y., Ali, M. A., Araki, S., Mwaura, J., & Nogami, Y. (2024). Optimizing IoT Intrusion Detection Using Balanced Class Distribution, Feature Selection, and Ensemble Machine Learning Techniques. *Sensors (Basel, Switzerland)*, 24. <https://doi.org/10.3390/s24134293>
16. Nassreddine, G., Nassreddine, M., & Al-Khatib, O. (2025). Ensemble Learning for Network Intrusion Detection Based on Correlation and Embedded Feature Selection Techniques. *Comput.*, 14, 82. <https://doi.org/10.3390/computers14030082>
17. Nigar, N., & Mustafa, R. (2025). Enhanced Intrusion Detection via Hybrid Data Resampling and Feature Optimization. *IEEE Access*, 13, 149100-149120. <https://doi.org/10.1109/access.2025.3602562>
18. Oh, S., Sohn, S., Kim, C., & Park, M. (2025). MCH-Ensemble: Minority Class Highlighting Ensemble Method for Class Imbalance in Network Intrusion Detection. *Applied Sciences*. <https://doi.org/10.3390/app152312647>
19. Ren, H., Tang, Y., Ren, S., Wang, R., & Dong, W. (2025). ADHS-EL: Dynamic ensemble learning with adversarial augmentation for accurate and robust network intrusion detection. *Journal of King Saud University Computer and Information Sciences*, 37. <https://doi.org/10.1007/s44443-025-00015-4>
20. Sălișteanu, I. C., Udriou, I., Gheorghe, A., Tudoroiu, I.-A., & Diaconu, E. (2026). Cost-Sensitive Stacking Ensemble with Hybrid Feature Selection for Rare Attack Detection in Network Intrusion Detection Systems. *Electronics*. <https://doi.org/10.3390/electronics15143094>
21. Shahiduzzaman, M., Sourav, M. S. A., Kabtia, M., Shammah, R. S., Islam, M. A., Chakraborty, P., Rahman, M. M., Islam, M., Tonoy, R. B. A., & Adnan, M. (2026). Attention-guided quantum-inspired extreme learning machine with grey wolf optimization for rare-attack detection in IIoT networks. *Discover Artificial Intelligence*. <https://doi.org/10.1007/s44163-026-01704-3>
22. Suresh, A., Sanjay, N., & Susan, S. (2026). Deepsmote-Driven Stacking Ensemble with Hybrid Feature Selection for Robust Multiclass Network Intrusion Detection. *2026 IEEE 18th International Conference on Computational Intelligence and Communication Networks (CICN)*, 711-717. <https://doi.org/10.1109/cicn70047.2026.11594261>

23. Talukder, M. A., Islam, M. M., Uddin, M. A., Hasan, K. F., Sharmin, S., Alyami, S., & Moni, M. A. (2024). Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. *Journal of Big Data*, 11, 1-44. <https://doi.org/10.1186/s40537-024-00886-w>
24. Upadhyay, L., Tripathi, M., & Grover, J. (2024). Performance Analysis of Ensemble Feature Selection based Network Intrusion Detection System using Ensemble Machine Learning. *2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, 1-6. <https://doi.org/10.1109/etncc63262.2024.10767475>
25. Yang, Z., Liu, Z., Zong, X., & Wang, G. (2022). An optimized adaptive ensemble model with feature selection for network intrusion detection. *Concurrency and Computation: Practice and Experience*, 35. <https://doi.org/10.1002/cpe.7529>